Image-based Detail Reconstruction of Non-Lambertian Surfaces

I-Chen Lin, Wen-Hsing Chang, Yung-Sheng Lo, Jen-Yu Peng, Chan-Yu Lin

Department of Computer Science,

National Chiao Tung University, Taiwan

Tel. (+886)3 5712121 ext 56684

email: ichenlin@cs.nctu.edu.tw

Abstract

This paper presents a novel optimization framework for estimating the static or dynamic surfaces with details. The proposed method uses dense depths from a structured-light system or sparse ones from motion capture as the initial positions, and exploits non-Lambertian reflectance models to approximate surface reflectance. Multi-stage shape-from-shading is then applied to optimize both shape geometry and reflectance properties. Because this method uses non-Lambertian properties, it can compensate for triangulation reconstruction errors caused by view-dependent reflections. This ap-

proach can also estimate detailed undulations on textureless regions, and employs spatial-temporal constraints for reliably tracking time-varying surfaces. Experiment results demonstrate that accurate and detailed 3D surfaces can be reconstructed from images acquired by off-the-shelf devices.

Keywords: Image-based 3D modeling, shape-from-shading, non-Lambertian reflection, motion capture, facial animation

Introduction

Recently, the demand for digitizing static or deformable surface parameters from real objects has increased dramatically. These digitized surfaces can be manipulated as 3D models in computer graphics and animation, or they can be used as statistic data in model-based vision or other analysis techniques.

Using images from one or multiple views is a flexible and effective approach to calculating 3D positions. Passive stereo triangulation which evaluates transformation and depths from pixel correspondences, is the most popular approaches to image-based modeling. However, pixel correspondences are usually ambiguous on textureless regions where estimated depths are also unreliable.

Structured-light reconstruction which uses a camera and a projector to acquire depth images, is another popular triangulation method. In 2004, L. Zhang et al. [1] introduced an impressive system for estimating dynamic face surfaces. They exploited space-time coherence to establish a more reliable pixel-correspondence. However, they had to use high resolution devices with a fast capturing speed due to the inherent properties of coded structured-light systems. Moreover, when we scan objects with non-lambertian properties, the projected light stripes can be distorted by specular or sub-surface scattering effects. For this reason, most experiments spread diffuse-reflected powders or paints on these kinds of objects.

On the other hand, photometric stereo methods and shape-from-shading (SFS) evaluate surface normals according to variations of image intensity. Most existing methods adopt

the Lambertian reflectance model. The photometric stereo calculates the surface normals from images under different light directions but from a fixed view point. However, using photometric stereo methods for dynamic objects require expensive high-speed cameras and a light dome [2][3]. Shape-from-shading methods usually illuminate the scene by a single light source and use only intensity gradients for shape recovery, avoiding ambiguity of pixel correspondence. The SFS methods are also sensitive to subtle variations and are therefore adequate for acquiring detailed undulation data. Nevertheless, its sensitivity to noise may lead to serious tremble.

The goal of this paper is to enhance the detail estimation capacity of existing 3D capture devices. The proposed technique focuses on structure-light reconstruction and facial motion capture systems, and can therefore be extended to other triangulation-based systems easily. The proposed method uses the shading information to amend or enrich 3D positions by stereo triangulation. In other words, stereo triangulation evaluates absolute positions around distinct features and SFS retrieves relative undulation.

For structured-light scanning systems, bidirectional reflectance distribution function (BRDF) and bidirectional surface scattering reflectance distribution function (BSSRDF) can both correct distorted or noise-contaminated surfaces. To capture facial motion, the proposed method uses optimized shape from shading to extract time-varying wrinkles and creases, which are difficult to capture using marker tracking. Furthermore, spatial-temporal coherences are employed for more reliable tracking. Figures 1 and 2 show the proposed method applied to range scanning and motion capture systems, respectively.

Several recent studies also share the idea of combining the advantages of positioning and shading information [4][5]. Bickel et al. [6] enhanced facial wrinkles using reflected intensities. In contrast, the proposed iterative optimization scheme uses only one image as an input for detailed shape recovery. Moreover, since we don't need to apply additional paints or restrict the shape of details, this method can even be applied to existing capture data or recorded images.

This paper is organized as follows. The following sections introduces most related research, followed by overview of the proposed method. The section titled "Estimating Surface Detail by SFS" proposes the progressive shape rectification method. The "Shape-from-shading for sparse motion data" section applies detailed recovery to a facial motion capture system. The final section concludes this paper.

Related Work

The proposed method estimates detailed surfaces using shape-from-shading based on graphics reflection models and space-time constraints. This section provides a literature review on these topics.

Stereo triangulation has been the main approach to vision-based shape recovery for decades. However, reliable pixel correspondence remains illusive in this approach. Readers may refer to D. A. Forsyth and J. Ponce [7] for a detailed explanation. In 2005, G. Vogiatzis et al. [8] proposed using the visual hull as the initial volume and optimized the

results by graphic-cut minimization. Their system provided a more reliable 3D surface but also suffered from the correspondence error caused by non-Lambertian reflection.

While applying triangulation for time-varying motions, conspicuous markers can alleviate this correspondence problem. B. Guenter et al. [9] proposed a dynamic face digitization system that tracked more than a hundred special markers. For time-varying facial details, they simply recorded dynamic texture. I.-C. Lin et al. [10][11] proposed capturing dense facial markers with mirror-reflected views. Even though they tracked nearly three hundred markers, they still had difficulty in reconstructing wrinkles or dimples. In 2008, U. Castellani et al. [12] proposed a robust deformable surface estimation method based on range scan data. They also accounted for spatial-temporal coherence, but excluded reflectance effect.

Anatomical models can mimic wrinkles and creases [13]. E. Sifakis, et al. [14] presented an automatic approach to determining muscle parameters based on sparse markers. However, this method cannot reproduce exact details of a performer.

On the other hand, photometric stereo and shape-from-shading methods evaluate delicate surface normals using variations of image intensity. Photometric stereo approaches usually analyze images from a fixed view point but under various lighting directions. Given accurately aligned pixels, photometric stereo methods can estimate surface normals by a simple least-square method. Empirically, more than eight images of different light directions are required to acquire reliable normals. A. Hertzmann and S.M. Seitz [15] proposed an example-based method in which the target surface normals were evaluated by comparing reference images of a known object. W.-C. Ma et al. [16] proposed a novel polarized illu-

mination method to acquire diffuse and specular normal maps. Their system is capable of estimating high-resolution detail face geometry.

In contrast, the shape-from-shading (SFS) method can avoid errors caused by inaccurate pixel correspondence, and this low-cost approach requires only one image for shape recovery. H. Fang and J.C. Hart [17] utilized a Lambertian-reflection-based method to extract the normal map from a single texture image. This kind of approach does not require expensive devices. However, it was error-prone due to the simple Lambertian assumption and was sensitive to input noise. As a result, manual adjustments were usually required in post processing.

SFS is also difficult to apply to real data due to its intrinsic ill-condition. In interactive modeling, G. Zeng et al. [18] proposed a semi-automatic solution for continuous surfaces. In this approach, users assigned surface normals to specific feature points and the system then refined the surface variations of the whole face. A. H. Ahmed and A. A. Farag [19] proposed using Oren and Nayar's diffuse reflection model for more accurate SFS calculations of rough surfaces. In a later study [20], they used perspective camera and Ward's reflection model. Since they proposed using a fast PDE solving method, their work cannot directly be applied to time-varying data with specific constraints, e.g. spatial-temporal coherence, or range constraints.

Recently, several researchers have combined stereo triangulation with shading information. D. Nehab et al. [4] proposed an efficient linear least square method that combines 3D positions with photometric normals. They assumed the objects have only lambertian re-

flectance. Yu et al [5] proposed a Phong-reflection-based shape recovery. They used visual hull of multiple views as the initial volume. Then, an optimization approach was employed to find the shape and reflectance parameters of static objects in the views.

B. Bickel et al. [6] proposed a multi-scale capturing system for facial motion. They first used conventional motion capture for large-scale motion. For middle-scale motions, they painted a specific color on each wrinkle and estimated parameters of the "valley-shape" wrinkle model from video.

Unlike most SFS under a Lambertian assumption, the proposed approach applies more general reflectance models. In computer graphics, the Bidirectional Reflectance Distribution Function (BRDF) is widely used to represent the reflectance model of human faces. Assuming that human skin is composed of an oil layer, epidermis, and dermis, skin reflectance can be approximated by a specular component at the oil-air interface and a diffuse reflectance component due to subsurface scattering. For more realistic subsurface scattering, Jensen et al. [21][22] introduced a Bidirectional Surface Scattering Reflectance Distribution Function (BSSRDF) model that combines dipole diffusion approximation and single scattering computation. Such sophisticated dipole models can also be embedded in our framework.

Overview

The approach presented in this paper combines the benefits of stereo triangulation and shape-from-shading methods. Stereo-based 3D reconstruction is employed to evaluate the

rough geometry, while a novel progressive shape-from-shading optimizes the detailed surfaces. We apply this framework to two applications: structured-light scanning and facial motion capture. For structured-light reconstruction, evaluated depths are set as the initial surface heights and range constraints, and the heights and reflectance parameters are refined progressively.

To capture facial motion, the proposed system first scattered the markers' 3D positions to non-marker regions by a radial-basis function (RBF). These scattered depths are then utilized as initial heights for surface optimization. In motion surface evaluation, surface heights are time-varying variables, and spatial and temporal constraints are included for more stable results. Figure 3 shows the flow chart of the proposed system.

Estimating Surface Detail by SFS

As mentioned above, stereo reconstruction can estimate accurate depths with conspicuous features, and structured-light systems that actively project light stripes can simplify the feature identification problems in textureless regions. We set up one video camera and one digital projector to form a structured light system. The projected light stripes are calibrated by a method proposed by D. Huynh [23]. Figure 4 shows the progress of scanning a marble statue.

Objective functions of SFS

After acquiring the initial positions, we utilize shape from shading to refine the surface according to shaded surface intensity. We adopt an optimization method that minimizes the difference between the captured image I and the synthesized image S . The objective function O_{static} becomes

$$O_{static} = (S - I)^2 \quad (1)$$

To decrease the degrees of freedom (DOF) of an objective function, without loss of generality, we represent the 3D data in terms of a height map. These heights can be easily transformed to normals from the gradients of neighbor z values. Hence, only the z (height) values of each aligned pixel (vertex) are evaluated. The shape parameters Z are therefore defined as:

$$Z = (z_1, z_2, \dots, z_p, \dots, z_{n_p})$$

where n_p is the number of pixels(vertices).

We choose the Phong model as the analytic reflection model, but other BRDF models can also be applied. The Phong model is widely used in computer graphics and its differentiation of parameters is relatively straightforward. Given a light source L with direction N_L and the surface normal N_p , the reflection intensity on vertex p related to normal can be written as:

$$S_p(N_p, K_d, K_s) = I_d \cdot k_d(N_L \cdot N_p) + I_s \cdot k_s(e \cdot r)^\alpha \quad (2)$$

where k_d and k_s are the diffuse and specular coefficients and α is the Phong exponent term.

The vector e denotes the view direction and r is the reflection vector with respect to N_L and N_p .

Assume that the reflectance parameter of a subject's face, $R=(k_d, k_s, \alpha)$, is uniform in a region. Equation (1) then becomes

$$O(Z, R) = \sum_{p=1}^{n_p} (S_p - I_p)^2 \quad (3)$$

In other words, our goal is to find the best surface sequence Z^* and reflectance parameter R^* that minimize the objective function. Instead of BRDF models, the BSSRDF model proposed by Jensen et al. [21] can also be applied as:

$$S_p(N_p, \sigma_s, \sigma_a) = \frac{\alpha z_r (1 + \sigma_{tr} d_{r,i}) e^{-\sigma_{tr} d_r}}{4\pi d_r^3} - \frac{\alpha z_v (1 + \sigma_{tr} d_{v,i}) e^{-\sigma_{tr} d_v}}{4\pi d_v^3} \quad (4)$$

where $\sigma_{tr} = \sqrt{3\sigma_a\sigma'_t}$ is the effective transport coefficient, $\sigma'_t = \sigma_a + \sigma'_s$ is the reduced extinction coefficient, $\alpha' = \sigma'_s/\sigma'_t$ is the reduced albedo, and σ_a and σ'_s are the absorption and reduced scattering coefficients. $z_r = 1/\sigma'_t$ and $z_v = (1 + 4A/3)/\sigma'_t$ are the z -coordinates of the positive and negative sources relative to the surface at $z=0$. $r = \|x_o - x_i\|_2$ and $d_r = \sqrt{r^2 + z_r^2}$ and $d_v = \sqrt{r^2 + z_v^2}$ are the distances to the sources from a given point on the surface of the object. A is defined as

$$A = (1 + F_{dr})/(1 - F_{dr})$$

where F_{dr} is the diffuse Fresnel reflectance. Readers can refer to [21] for a detailed explanation.

Multiple-stage optimization

Experiment results show that the estimated heights were highly sensitive to noise when applying direct shape-from-shading methods [17]. When we used the heights from structured-light scanning as initial values and applied modified conjugate-gradient methods, these variables were easily trapped into local minima in early stages, producing undesired trembling effects. For more stable estimation, we propose using a multi-stage optimization scheme.

Our objective function has two sets of parameters: the shape parameter Z and reflectance parameter R . If we simultaneously estimate these two parameters into a single stage, we need a proper scale between these two kinds of parameters for balanced influences. To avoid bias, we optimize these two sets of parameters separately.

Experiment results also show that when directly apply all BRDF parameters in optimization, the specular components dominated the objective function, trapping the process in a local minimum. On the other hand, the reflections of human skin or marble statues are mainly contributed by diffuse components. Hence, we introduce an intermediate stage, diffuse-shape optimization in which the shape Z and diffuse parameter $R_{diffuse}$ (kd) are iteratively optimized. After this stage, BRDF-shape optimization is performed for all diffuse and specular parameters. Figure 5 shows the flow chart of the multi-stage optimization. Figure 6 shows a conceptual diagram of optimized shape-from-shading.

Spatial coherence and range constraints

Since our targets, e.g faces, are mostly continuous surfaces, the motion of a vertex p has a high spatial coherence with its neighbors. Therefore, we use spatial coherence constraints to alleviate the inherent noise sensitivity in the shape-from-shading approach .

$$CS_p = k_{CS}(z_p - \sum_j \frac{1}{w_j} z_j)^2, for j \in Neighbor(tp) \quad (5)$$

, where $Neighbor(p)$ denotes the 8-neighbor pixel set, w_j is an adaptive weight, and k_{CS} is the weight for spatial constraints. After transforming the corresponding height-map block to the frequency domain, an area will be more influenced by its neighbors if it is dominated by low-frequency components beyond a certain threshold. Figure 7 shows the influence of spatial constraints.

Moreover, since the scanned surfaces are near the exact positions and SFS is aimed at adjusting relative undulation, range constraints CR can be used to keep the heights near the original scans.

$$CR_p = -k_{CR}e^{-(z_p - z'_p)^2} \quad (6)$$

where k_{CR} is the weight for range constraints and z'_p is the estimated height of pixel p by triangulation.

Shape-from-shading for sparse motion data

The space-time study of L.Zhang et al. [1] extends a stereo structured-light system to facial motion capture. Nevertheless, structured-light reconstruction acquires multiple stripe sets for one surface, and therefore requires higher capture rate for time-varying surfaces.

Instead, the proposed study tracks facial details by extending widely-used motion capture techniques. Since motion capture techniques only track markers' 3D trajectories, we first have to reconstruct the rough initial surfaces according to the sparse markers' 3D positions.

Initial surfaces from motion capture

To evaluate time-varying primitive 3D face surfaces, we adopt a model-based approach. First, the 3D positions of markers in the first frame are estimated by stereo triangulation. We then characterize a generic face model by feature-point fitting, which is similar to the deformation method described below.

Assume that the expression at the initial frame is neutral. For each of the following frames, track the markers and deform the characterized model according to the markers' 3D positions. The deformation method employed in this study is radial-basis-function-based (RBF-based) data scattering.

Consider a set of corresponding pairs $\{q_i, q(t)_i\}$ between the neutral face and an expressional face at time t , where q_i is the 3D position of the marker i on the neutral face, and

$q(t)_i$ is the position of the marker on an expressional face at time t . We define the motion (displacement) of each marker as $u(t)_i = q(t)_i - q_i$, and use scattering function $F(t, q)$ to estimate the displacement of a non-feature point. The scattering function at time t is then

$$F(t, q) = \sum_i c(t)_i \varphi(\|q - q_i\|) + A(t)q + B(t) \quad (7)$$

where $\varphi(r) = e^{-r/32}$ is a radial basis function. $c(t)_i$ are weighted coefficients, and $A(t)$, $B(t)$ are affine terms. The coefficients $c(t)_i$, $A(t)$ and $B(t)$ can easily be solved by linear equations with the following constraints:

$$u(t)_i = F(t, q_i), \sum_i c(t)_i = 0, \sum_i c(t)_i q_i = 0$$

Spatial-temporal coherence

When applying our optimized SFS to each frame individually, we found that there were still flickers caused by input noise. According to the biomechanical properties of facial muscles and tissues, the surface of a human face should gradually transfer between expressions. Hence, temporal coherence can further improve stability.

The first step is to extend our shape representation Z to a time-varying form:

$$Z_t = (z_{t1}, z_{t2}, \dots, z_{tp}, \dots, z_{t(n_p)})$$

where t is an index of time (or frame).

Applying optimization to the entire image sequence is extremely time-consuming. For computational efficiency, we only apply our optimization to a small window of frames at a

time (7 frames in this case). A pseudo-optimal surface sequence can be acquired by sweeping the windows from the start to the end and combining the individual results. This approach significantly decreases the degrees of freedom and dramatically reduces processing time.

The temporal constraint then becomes

$$CT_{tp} = k_{CT}(z_{tp} - \sum_i \frac{1}{w_i} z_{(t+1)p})^2, \text{ where } i = [-3, 3] \quad (8)$$

Therefore, including time-varying variables and spatial-temporal coherence and range constraints, our objective function becomes

$$O = \sum_{t=1}^{n_t} \sum_{p=1}^{n_p} [(S_{tp} - I_{tp})^2 + CS_{tp} + CT_{tp} + CR_{tp}] \quad (9)$$

where S_{tp} , I_{tp} , CS_{tp} , and CR_{tp} are the extensions of S_p , I_p , CS_p and CR_p with time indices.

Experiments and results

This section describes and discusses our experiments and error analysis, and presents results of static objects and facial motions.

Experiments

The proposed system uses one video camera and one projector for structured-light reconstruction. To simplify the optimization, all our experiments were performed under an illumination-controlled environment. A calibrated spotlight served as a single directional light source.

The high-definition video(HDV) cameras used in this setup had a resolution of 1280*720 pixels and a 30 frames-per-second capture rate.

For facial surfaces, two synchronized video cameras were applied to track marker motions. We pasted 25-30 markers on subjects' faces, being careful not to place markers on regions with wrinkles or creases. Three subjects, one female and two males, were requested to give various facial expressions. Since the proposed technique is designed to enhance delicate variations in motion capture, users can assign regions for detail estimation. In our experiments, we preferred areas with more creases, such as the forehead, glabella, left and right cheeks. Figure 9 shows the captured neutral image and the user-assigned regions by a brush-based interface.

Moreover, since our goal is to extend the usage of existing devices, the proposed system can use the images from structured-light or motion capture to optimize SFS directly. In motion capture, we used two views, and therefore two height maps were acquired. Given two synchronized height maps and feature points, the view morphing technique [24] was applied to warp the two side views into a single front one. Then, we weighted-blended the two sets of view data according to the view angles. Data that are closer to their original view get a higher ratio.

The proposed sytem adopts a conjugate-gradient-based method, called the Broydon-Fletcger-Goldfarb-Shanno (BFGS) method, for non-linear function minimization. At each iteration, we first estimated the approximate step direction, and the golden section search was further utilized to find the optimal step size. In facial motion estimation, where four

mid-small size regions are evaluated, the performance of our optimization was around 120 to 200 seconds per frame on a PC with a 3.2GHz CPU; for scanned objects, the minimization took nearly one hour per view. Since the current system only uses single-threaded programming, the computation time can be further improved with multi-threaded programming.

Figure 10 demonstrates the progress of cost versus iteration time for three exemplary frames in a facial motion sequence. The cost here is the error between the input image and synthesized image. In our evaluation, we normalized the intensity of a pixel to range $[0,1]$. The cost gradually decreased during the diffuse-shape optimization. When specular parameters were included in reflectance optimization, the cost dropped significantly. BRDF-shape optimization further improved the shape and evaluated BRDF parameters.

Error analysis

We evaluated the performance of the proposed detail estimation method by computer simulation. Two virtual models, "bunny" and "human head", were selected as targets. We assumed the "heights" ($\max y - \min y$ values) of these models were 200mm, and applied random noises of maximum 2mm to their vertices. Given the noise-contaminated models and the shaded images of ground-truth models, we applied our multi-stage SFS for surface rectification. Figure 11 shows the experiments based on Phong reflection model. In this case, the per-vertex errors of the human head were reduced from 0.634715mm to 0.236203mm, while the bunny's were reduced from 1.035466mm to 0.991826mm. Even though the error

improvement of the high curvature "bunny" model is less obvious, the refined model is visually more appealing and closer to the ground-truth due to our appearance-based algorithm.

We also performed this analysis to BSSRDF-reflection objects. Nevertheless, we found that the optimization of BSSRDF parameters diverged easily and required precise initial guesses. Therefore, in BSSRDF experiments, we only performed height optimization and the BSSRDF reflectance parameters were assumed to be known. Figure 12 shows that random noises of maximum 5mm were embedded in the bunny model. After our position optimization with BSSRDF reflection, the average per-vertex error was reduced from 3.554mm to 0.031mm.

Figure 13 shows the progress of optimization in terms of stages. In the first stage, the diffuse term is optimized to obtain a more accurate shape. The second stage includes specular terms. The estimated surfaces gradually approached the input image sequence. Figure 14 also compares our space-time SFS with a direct Lambertian-based method [17]. Our method provides more accurate surface and suppresses noise.

Synthesis of Faces

We derived a characterized neutral face according to motion capture data. Our 3D head model contained 6078 vertices and 6315 polygons. For deformation, we applied RBF-based data scattering to each segment and gradually blended the boundaries. To enhance details with estimated height maps, the target polygons were first subdivided and per-pixel normal

mapping was then applied. Figure 15 depicts the face with detail wrinkles. The synthesized results can be further improved with relief textures and realistic face rendering. Besides, our acquired motion can be also employed for detailed expression synthesis, e.g. the work of Golovinskiy et al. [25].

Discussion and conclusion

The goal of this paper is to estimate non-Lambertian-reflected surface details by off-the-shelf devices. The proposed approach combines the advantages of stereo triangulation and shape-from-shading for static scanning and facial motion. The depths estimated from triangulation are employed as initial surfaces and range constraints for optimization. The system then minimizes the error between the input and synthesized images.

To tackle the optimization problem with considerable reflectance and time-varying shape parameters, we propose an iterative scheme with multiple stages, in which diffuse parameters, diffuse-specular parameters, and the heights of surfaces are improved sequentially. Spatial and temporal coherence constraints are also included for more reliable estimation. Simulations and experimental results demonstrate the effectiveness of the proposed method.

Compared with the related articles [5][6], where impressive shapes were evaluated from multiple static views or specific groove models, the proposed method is more versatile for various details with dense or sparse initial depths. We neither assume the shape of details nor spread powder or paint for diffuse reflection. If the illumination is controlled, the proposed

method can even be applied to existing scanning or motion capture video.

The proposed method can be adapted to various non-lambertian reflection models. To our best knowledge, the proposed system could be the first article applying the BSSRDF reflection [21] to shape from shading. Experiment results show that a precise evaluation of reflectance parameters significantly improve the shape recovery, but optimizing reflection parameters is easily trapped in local minima, especially for complex subsurface scattering. Therefore, we can only perform position optimization in that situation. For real BSSRDF surface estimation, more intensive computations with global or heuristic optimization, e.g. simulated annealing, can alleviate this situation. The current approach assumes that reflection properties are uniform at each user-defined region. We plan to include automatic segmentation of the surface reflection in the future.

References

- [1] L.Zhang, N.Snively, B.Curless, and S.M.Seitz. Spacetime faces: High resolution capture for modeling and animation. *Proceedings of ACM SIGGRAPH*, pages 548–558, 2004.
- [2] A.Wenger, A.Gardner, C.Tchou, J.Unger, T.Hawkins, and P.Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graphics*, 24(3):756–764, 2005.

- [3] T.Weyrich, W.Matusik, H.Pfister, B.Bickel, C.Donner, C.Tu, J.McAndless, J.Lee, A.Ngan, H.W.Jensen, and M.Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graphics*, 25(3):1013–1024, 2006.
- [4] D.Nehab, S.Rusinkiewicz, J.Davis, and R.Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *Proceedings of ACM SIGGRAPH*, pages 536–543, 2005.
- [5] T.Yu, N.Xu, and N.Ahuja. Recovering shape and reflectance model of non-lambertian objects from multiple views. *Proc. IEEE CVPR*, pages 226–233, 2004.
- [6] B.Bickel, M.Botsch, R.Angst, W.Matusik, M.Otaduy, H.Pfister, and M.Gross. Multi-scale capture of facial geometry and motion. *ACM Trans. Graphics*, 26(3):33.1–33.10, 2007.
- [7] D. A. Forsyth and J. Ponce. Computer vision: A modern approach. *Prentice Hall*, 2002.
- [8] G.Vogiatzis, P.H.S.Torr, and R.Cipolla. Multi-view stereo via volumetric graph-cuts. *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2:391–398, 2005.
- [9] B.Guenter, C.Grimm, D.Wood, H.Malvar, and F.Pighin. Making faces. *Proceedings of ACM SIGGRAPH*, pages 55–66, 1998.
- [10] I.-C.Lin, J.-S.Yeh, and M.Ouhyoung. Extracting 3d facial animation parameters from multiview video clips. *IEEE Computer Graphics and Applications*, 22(6):72–80, 2002.

- [11] I.-C.Lin and M.Ouhyoung. Automatic and efficient capture of dense 3d facial motion parameters from video. *The Visual Computer*, 12(6):355–372, 2005.
- [12] U.Castellani, V.Gay-Bellile, and A.Bartoli. Robust deformation capture from temporal range data for surface rendering. *Computer Animation and Virtual Worlds*, 19:591–603, 2008.
- [13] N.Magnenat-Thalmann, P.Kalra, J.Luc Leveque, R.Bazin, D.Batisse, and B.Querleux. A computational skin model: Fold and wrinkle formation. *IEEE Trans. Information Technology in Biomedicine*, 6(4):317–323, 2002.
- [14] E.Sifakis, I.Neverov, and R.Fedkiw. Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Trans. Graphics*, 24(3):417–425, 2005.
- [15] A.Hertzmann and S.M.Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(8):1254–1264, 2005.
- [16] W.Ma, T.Hawkins, P.Peers, C.Chabert, M.Weiss, and P.Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Eurographics Symposium on Rendering*, 2007.
- [17] H.Fang and J.C.Hart. Textureshop: Texture synthesis as a photograph editing tool. *ACM Trans. Graphics*, 23(3):354–359, 2004.

- [18] G.Zeng, Y.Matsushita, L.Quan, and H.Y.Shum. Interactive shape from shading. *Proceedings of IEEE Conf.Computer Vision and Pattern Recognition*, 1:343–350, 2005.
- [19] A.H.Ahmed and A.A.Farag. A new formulation for shape from shading for non-lambertian surfaces. *Proceedings of IEEE Conf.Computer Vision and Pattern Recognition*, 2:1817–1824, 2006.
- [20] A.H.Ahmed and A.A.Farag. Shape from shading under various imaging conditions. *Proceedings of IEEE Conf.Computer Vision and Pattern Recognition*, pages X1–X8, 2007.
- [21] H.W.Jensen, S.Marschner, M.Levoy, and P.Hanrahan. A practical model for subsurface light transport. *Proceedings of ACM SIGGRAPH*, pages 511–518, 2001.
- [22] C.Donner and H.W.Jensen. Light diffusion in multilayered translucent materials. *ACM Trans. Graphics*, 24(3):1032–1039, 2005.
- [23] D.Huynh. Calibration of a structured light system: A projective approach. *Proc. IEEE CVPR*, pages 225–230, 1997.
- [24] S.M.Seitz and C.R.Dyer. View morphing. *Proceedings of ACM SIGGRAPH*, pages 21–30, 1996.
- [25] A.Golovinskiy, W.Matusik, H.Pfister, S.Rusinkiewicz, and T.Funkhouser. A statistical model for synthesis of detailed facial geometry. *ACM Trans. Graphics*, 25(3):1025–1034, 2006.

Figure 1: The proposed optimized shape-from-shading for detail refinement. (a). A captured image of the target statue. (b). The estimated surface by structured-light triangulation. The depths are distorted by input noise and stripe calibration errors. (c) The surfaces estimated by the proposed optimization. (d) The synthesized model with texture mapping.

Figure 2: The proposed method for time-varying facial detailed motion. (a) A captured image of facial motion capture. (b) The retargeted face without facial details. (c) The synthesized face enhanced by the estimated wrinkles.

Figure 3: The flow chart of the proposed system. The optimization for static objects is in blue and that for facial motions is in green.

Figure 4: Scanning a marble statue by a structured-light system.

Figure 5: The flow chart of multi-stage optimization.

Figure 6: A conceptual diagram of progressive surface estimation with a height map.

Figure 7: The influence of spatial coherence constraints. (left) the optimized surface without spatial constraints. (right) the result with spatial constraints.

Figure 8: Deformation of a generic model for personalization and facial expression. (left) the generic model; (middle) the personalized neutral face; (right) eyebrow raising according to markers' positions.

Figure 9: The captured two-view image and extracted detailed surfaces. (a)(d) are the right and left views of the neutral face; (b)(c) are the user-assigned regions for evaluation; (e)(h) are the right and left views of an expressional face; (f)(g) are the estimated detailed surfaces of (e) and (h).

Figure 10: Iterative recovery of a facial surface sequence. The leftmost column shows the input image; The middle two columns demonstrate the estimated normal maps at diffuse-shape and BRDF-shape stages. The rightmost column shows the cost versus iteration time.

Figure 11: Error analysis of SFS correction with Phong-reflection objects by simulation.

(a) The ground-truth bunny; (b) the noise-contaminated bunny; (c) the SFS-corrected bunny;
 (d) the ground-truth head; (e) the noise-contaminated head; (f) the SFS-corrected head.

Figure 12: Error analysis of SFS correction with BSSRDF-reflection objects (with known BSSRDF- reflection parameters). (a)The ground-truth bunny; (b) the noise-contaminated bunny; (c) the SFS-corrected bunny.

Figure 13: The progressive surface refinement for a captured wrinkle image.

Figure 14: Comparison of the proposed space-time SFS method and a Lambertian-based direct SFS method.

Figure 15: Facial details synthesized by subdivision and normal mapping.